using the dot blot assay described here. This assay has been successfully used in several laboratories.[16,17]

Acknowledgment

The author thanks R. Reeder for helpful suggestions on this manuscript. This work was supported by a grant from the National Institutes of Health to Harold Weintraub. P. S. T. is a Postdoctoral Fellow of The National Institutes of Health.

[16] M. Groudine, R. Eisenman, and H. Weintraub, *Nature* (*London*) **292**, 311 (1981).
[17] S. F. Wolf and B. R. Migeon, *Nature* (*London*) **295**, 667 (1982).

[19] Isolation of Multigene Families and Determination of Homologies by Filter Hybridization Methods

By GERALD A. BELTZ, KENNETH A. JACOBS, THOMAS H. EICKBUSH, PETER T. CHERBAS, and FOTIS C. KAFATOS

A high proportion of genes in eukaryotes are now known to be members of multigene families. Although initial indications were provided by protein sequencing, the ubiquity of multigene families has been revealed most convincingly by recombinant DNA methods. A list of some of the best known multigene families, by no means inclusive, would include the genes for histones, globins, immunoglobulins, histocompatibility antigens, actins, tubulins, seed storage proteins, chorion proteins, keratins, collagens, cuticle proteins, yolk proteins, heat shock proteins, and salivary glue proteins. Even genes for proteins that appear homogeneous may belong to multigene families; e.g., the chicken ovalbumin gene is now known to be expressed coordinately with two homologous genes, *X* and *Y*. In the field of hormone research, analysis of gene families related to well known effectors such as insulin or growth hormone is one of the most exciting areas of current investigation. Many members of multigene families show stage- or tissue-specific expression: a familiar example is the family of globin genes. Thus, the study of multigene families is important for understanding the mechanisms of physiological regulation, differential gene expression, and molecular evolution in eukaryotes.

In the study of multigene families by recombinant DNA methods, usually the first step is to isolate clones (cDNA or genomic) that represent all members of the family. The next step is to discriminate among clones that correspond to different genes. The third step is to relate each gene to

ISBN 0-12-182000-9

a particular type of protein product or to a specific RNA transcript. Ultimately, structural characterization of the genes, including sequence analysis, is undertaken. Because of their sequence homologies, members of a multigene family often cross-hybridize. In the first three steps of such studies, careful control of the hybridization conditions is necessary—both to isolate the family members as cross-hybridizing species in the first step, and to permit their discrimination and characterization in the second and third steps. A number of filter hybridization methods are used in these studies: phage plaque[1] and bacterial colony hybridization,[2] dot blot hybridization,[3] Southern hybridization,[4] Northern hybridization,[5] and hybrid-selected translation.[6] Their specific features are discussed elsewhere in this volume. Here, we wish to emphasize their relatedness and to point out how deliberate control of the hybridization stringency maximizes their utility. We shall give examples from our studies of the chorion gene families in silkmoths.

The silkmoth chorion system has been reviewed.[7] More than a hundred genes belonging to several multigene families encode the structural proteins of the chorion. These families may themselves be related, constituting a superfamily,[8] but the interfamily homologies are low and can be neglected in this discussion. Within each family, both closely and distantly related genes are found: observed degrees of mismatching range from less than 1% to as much as 50%.[9] Based on this variation, members of each family have been classified into distinct *types* (distantly related genes, observed mismatching 10 to 50%) and into *copies* of the same gene type (more closely related genes, <1 to 5% mismatching). In our experience, even gene copies differ by one or more substitutions that lead to amino acid replacements, and thus have already taken the first step toward evolving into distinct genes, by the criterion of differences in the encoded proteins. The sequence variations are not uniformly distributed within the chorion genes: in each family the sequence corresponding to a central protein domain is substantially conserved, whereas flanking se-

[1] W. D. Benton and R. W. Davis, *Science* **196**, 180 (1977).
[2] M. Grunstein and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961 (1975).
[3] F. C. Kafatos, C. W. Jones, and A. Efstratiadis *Nucleic Acids Res.* **7**, 1541 (1979).
[4] E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
[5] J. C. Alwine, D. J. Kemp, and G. R. Stark, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5350 (1977).
[6] R. P. Ricciardi, J. S. Miller, and B. E. Roberts, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4927 (1979).
[7] F. C. Kafatos, *Am. Zool.* **21**, 707 (1981).
[8] J. C. Regier, F. C. Kafatos, and S. J. Hamodrakas, *Proc. Natl. Acad. Sci. U.S.A.* (in press).
[9] C. W. Jones and F. C. Kafatos, *J. Mol. Evol.* **19**, 87 (1982).

quences that encode protein "arms" (NH_2-terminal and COOH-terminal domains) are much more variable.[10]

Parameters Affecting Nucleic Acid Hybridizations

A Qualitative Summary

By definition, the course of any hybridization reaction is determined by the concentrations of the reacting species and the second-order rate constant k. The stability of the resulting duplex is measured by the melting temperature, T_m. For reassociation reactions in solution involving perfectly matched complementary strands, the effects of various reaction conditions have been investigated in detail, and thus T_m and k values can be calculated with some precision. Neither calculation can be extended with confidence to cases involving imperfect hybrids or to solid-phase reactions. Nevertheless, the effects of reaction conditions can be expected to be qualitatively similar, and thus they bear review here.

In general, T_m depends on the composition of the duplex (T_m increases with G + C content), on the duplex length (T_m increases with length), on the ionic strength (T_m increases with salt concentration up to a plateau), and on the concentration of any organic denaturants present (e.g., formamide, urea). T_m is higher for RNA : DNA hybrids than for the corresponding DNA : DNA duplexes, and the differential effect of increasing (G + C) is greater for the hybrids.

Parameters such as (G + C), duplex length, ionic strength, and concentration of any denaturants generally affect k in the same direction as they affect T_m. Therefore, an additional parameter, which is of paramount importance for k, is the temperature of the reaction, T_i. In general, k increases with T_i, reaches a broad maximum at 25 to 20° below T_m, and decreases thereafter, becoming severely depressed at $T_i \geq T_m - 5°$. An example of such temperature dependence is shown in Fig. 1, for the reassociation of T4 DNA in solution.[11]

The qualitatively similar effects of most reaction conditions on k and T_m have led to the practice of summarizing the reaction conditions in terms of the *criterion,* which is numerically equal to T_m minus T_i. When the criterion is large, the reaction is described as permissive; when it is small, the reaction is stringent. While the *absolute* values of k vary drastically as a function of individual parameters (e.g., salt), the *relative* values as a function of criterion are thought to be described in general by the

[10] S. J. Hamodrakas, C. W. Jones, and F. C. Kafatos, *Biochim. Biophys. Acta* **700**, 42 (1982).

[11] T. I. Bonner, D. J. Brenner, B. R. Neufeld, and R. J. Britten, *J. Mol. Biol.* **81**, 123 (1973).

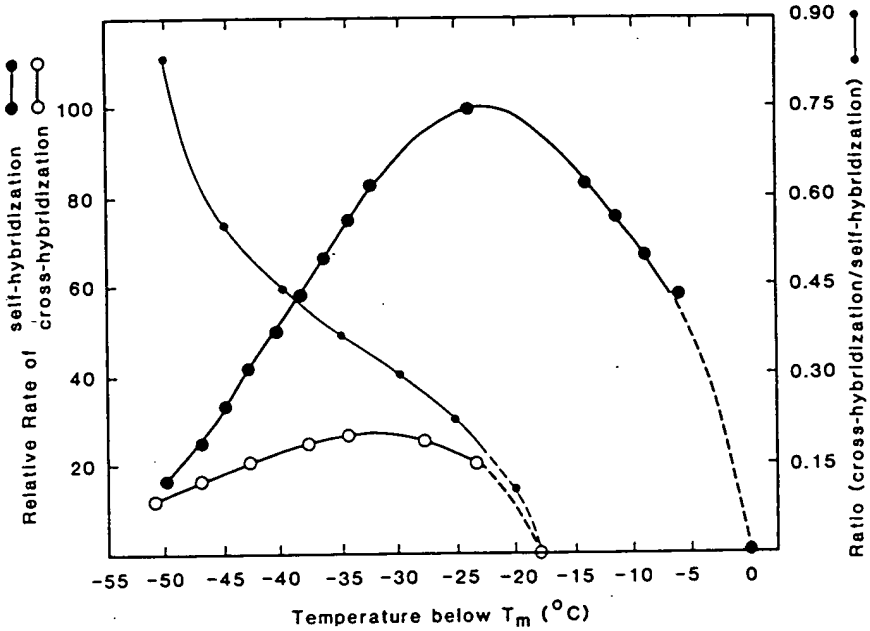


FIG. 1. Rate of reassociation as a function of temperature for normal (self-hybridization) and mismatched DNA (cross-hybridization). Data are replotted from Fig. 3 of Bonner *et al.*[11]; they are derived from normal T4 DNA and T4 DNA partially deaminated with nitrous acid. The large open and filled circles are from the best-fit curve plotted in the original figure, and dotted lines are extrapolations assuming that at T_m the rate of reassociation is zero. Under the conditions of the experiment (0.12 *M* phosphate buffer), the T_m of normal T4 DNA is 81°. The discrimination ratio, i.e., k of cross-hybridization divided by k of self-hybridization, is also presented.

curve of Fig. 1, for the reassociation of perfectly matched strands in solution.[11]

The problem of discrimination among related nucleic acid sequences is one of devising hybridization conditions that are stringent for some sequences and permissive for others. Although it is known that mismatching depresses both T_m and k, its effect on the relationship between k and criterion has not been studied in detail. Figure 1 presents one of the few sets of systematic data available for hybridization between normal T4 DNA and T4 DNA partially deaminated with nitrous acid (resulting in purine–pyrimidine mismatches that lower the T_m value by 18°). The temperature dependence of k for this "cross-hybridization" (T4 DNA vs deaminated T4 DNA) shows some similarity to that for "self-hybridization" (T4 DNA vs T4 DNA), but the rate constant is lower and reaches its optimal value at a lower temperature.

From data such as those of Fig. 1, one may calculate what we will call the "distribution ratio" for each temperature, i.e., the ratio between the rate constants for cross-hybridization and for self-hybridization reactions. The higher the value of this ratio, the easier it is to detect even distant homologs; the lower the ratio, the easier it is to discriminate between self-hybrids and cross-hybrids. It can be seen that the ratio varies with criterion in a sigmoidal manner. Although the exact shape of the sigmoid is not based on extensive empirical data, its general shape is not affected by minor variations in the individual k vs temperature curves: it depends on the bell shape of the curves and the shift of the cross-hybridization curve to lower temperatures.

The sigmoidal curve for discrimination ratio vs temperature has important implications for the study of multigene families. If we imagine a set of such curves, corresponding to different degrees of mismatching, we can understand why it is not possible to find a compromise criterion that would permit both detection of distant homologs and discrimination between close homologs. If both goals are to be attained, the following two-step procedure is necessary.

Step 1. To recover all members of a multigene family, an initial screen should be performed at a very permissive criterion. From Fig. 1, it can be seen that a criterion 40° to 50° below T_m is desirable if very distant homologs are to be recovered: at these temperatures the rate of self-hybridization is reduced two- to six-fold, but even distant homologs give a relatively strong signal. A practical limit is imposed by background problems. For example, if the probes have been cloned by G : C tailing, nonspecific hybridization due to the tails becomes significant at very permissive criteria.

Step 2. To discriminate between homologs, a very stringent criterion should be used in a second hybridization step. Since it appears that the discrimination ratio drops steeply near the T_m of the mismatched hybrids (Fig. 1), we recommend the use of criteria near the T_m of the relevant cross-hybrids. Temperatures as high as 5° below the T_m of the self-hybrids can be used without difficulty.

In theory, these two steps might be compressed into a single two-step experiment: After an initial hybridization under permissive conditions, poorly matched hybrids might be recognized by their release in a subsequent wash under stringent conditions. In practice, although this protocol is useful for discriminating between distant homologs,[3] we have found it not to be useful when the homologs are closely related. In part, the problem is technical. Although probes are for the most part reversibly hybridized, in filter hybridization a variable portion of the probe may become irreversibly bound. This problem can be minimized by scrupu-

lously avoiding drying the filter until after the melting reaction, but the precaution is not always sufficient.

A more important consideration affecting the design of the experiment is that a stringent hybridization and a stringent wash are not entirely equivalent. *A priori* the stability of a hybrid during washing depends on the T_m of the best-matched region of duplex. By contrast k reflects the nucleation frequency, which is an unknown function of the distribution of matched and unmatched stretches. In practice we find that a stringent hybridization is more discriminatory than an equally stringent wash.

Quantitative Considerations

It is clear from the preceding discussion that in the study of multigene families, selection of appropriate conditions requires attention to the parameters that govern T_m and k. We emphasize that systematic empirical data for matrix-supported reactions, and for mismatched sequences, are extremely limited. Thus, the appropriate conditions must be determined empirically in each case. However a starting point can be selected by calculations, and we summarize here the pertinent information from the literature for convenience in doing so.

The rate constant for renaturation of randomly sheared DNA in solution is given by the Wetmur–Davidson equation[12]

$$k = (k_N L^{0.5})/N$$

where L is the mean length of the reassociated duplex per nucleation; N is the complexity of the DNA; and k_N is a length-independent nucleation rate constant. The length factor is quite uncertain, because it attempts to incorporate effects of length on both nucleation frequency and yield of duplex per nucleation. Therefore, the equation can be used only for a rough estimate if the hybridizing strands differ in length. The constant k_N is influenced by such environmental factors as the incubation temperature, the salt concentration, organic solvents, and various polymers.[13] In an aqueous solution at an incubation temperature of $T_m - 25°$ and 1 M Na^+, $k_N = 3 \times 10^5$ liters/mole-second. Attachment of one of the reacting species to a matrix decreases k_N, probably by an order of magnitude.[13a] In addition, if a large amount of DNA is filter-bound, the hybridization may be limited by diffusion of the probe to the filter, unless the reaction vessel is shaken.[13a]

The effect of individual parameters on k is as follows. The change in

[12] J. G. Wetmur and N. Davidson, *J. Mol. Biol.* **31**, 349 (1968).
[13] R. Wieder and J. G. Wetmur, *Biopolymers* **20**, 1537 (1981).
[13a] R. A. Flavell, E. J. Birfelder, J. P. Sanders, and P. Borst, *Eur. J. Biochem.* **47**, 535 (1974).

optimal k at $T_m - 25°$ as a function of salt has been tabulated,[12,14] and is most dramatic below 0.4 M Na^+. An 1% increase in G + C increases[12] optimal k by a factor of approximately 0.018. Mismatching is reported[11,14] to depress the optimal k by a factor of 2 per 10° reduction in T_m. Concentrations of 80% formamide are thought to depress k by a factor of 3 for DNA–DNA hybrids and by a factor of twelve for RNA–DNA hybrids, although the data are limited.[15] Volume-excluding inert polymers can be used to increase k.[13,15a] In an aqueous solution of 1 M NaCl at 70°, the renaturation rate increases 100-fold when dextran sulfate is added to 40%.[13]

Concerning T_m, a large number of quantitative studies can be summarized as follows[11,15,16]: For a perfectly matched DNA duplex,

$$T_m = 81.5 + 0.41\,(G + C) + 16.6 \log(Na^+) - 0.63(\%\ \text{formamide}) - \frac{300 + 2000(Na^+)}{d}$$

where G + C = percentage of guanine + cytosine; Na^+ = molarity of (Na^+) or equivalent monovalent cation; and d = the length of the hybridized duplex in nucleotides.

It is prudent to bear in mind the limits of each term in this formula. The dependence on G + C is accurate over the range 30–75% G + C.[17] The salt dependence[16] is valid for 0.01 to 0.40 M (Na^+), only approximately so at higher (Na^+); T_m is maximal at about 1.0–2.0 M (Na^+). The depression of T_m by formamide[15] is greater for poly(dA : dT) (0.75°/1% formamide) than for poly(dG : dC) (0.50°/1% formamide), and the value used in the formula was derived from human rDNA (58–67% G + C). The effect of polynucleotide length on T_m is somewhat controversial, but the correction listed[14] seems valid over the range 0.05–0.5 M salt. Finally, the formula applies to the "reversible" T_m, e.g., as assayed by optical measurements. The "irreversible" T_m, which is of relevance for autoradiographic detection, is usually higher by 7–10° in aqueous solutions.[18]

For estimating the T_m of RNA–DNA hybrids, the paper of Casey and Davidson[15] should be consulted. The effect of formamide on such hybrids is nonlinear (unlike the effect on DNA–DNA hybrids), leading to the well-known preferential formation of RNA–DNA hybrids at high formamide concentrations. At 80% formamide, 0.3 M Na^+, RNA–DNA hybrids are

[14] R. J. Britten, D. E. Graham, and B. R. Neufeld, this series, Vol. 29, p. 363.
[15] J. Casey and N. Davidson, *Nucleic Acids Res.* 4, 1539 (1977).
[15a] G. M. Wall, M. Stern, and G. R. Stark, *Proc. Natl. Acad. Sci. U.S.A.* 76, 3683 (1979).
[16] C. Schildkraut and S. Lifson, *Biopolymers* 3, 195 (1965).
[17] J. Marmur and P. Doty, *J. Mol. Biol.* 5, 109 (1962).
[18] K. Hamaguchi and E. P. Geiduschek, *J. Am. Chem. Soc.* 84, 1329 (1962).

reported to be 20–30° more stable than DNA–DNA hybrids. At 50% formamide, 0.3 M Na^+, we find that the difference is approximately 11° in the case of β-globin sequences.[3]

The effect of mismatching on T_m is obviously very important for studies on multigene families. The usual simplification is that T_m decreases by 1° for every 1 ± 0.3% mismatch,[14] reckoned for DNA–DNA duplexes of 40% G + C with randomly placed mismatches.[14] In another study, percentage homologies for BK and SV40 DNA fragments determined by heteroduplex analysis gave the best fit to actual sequence data when 0.5%/1% mismatch was used.[18a] Obviously the distribution of mismatched bases will have a critical determining effect. Consider an extreme example: sequence A differs from sequence B by 50%. If the mismatch is clustered so that extended regions of A and B are identical is sequence, a high T_m will be observed; by contrast, if the mismatch is dispersed so that every second nucleotide differs, no hybridization will occur. The effect of sequence on the hybridization of short oligonucleotides with mismatches has been studied.[19,20] The extrapolation of these very interesting results to real-life problems will require some judgment and additional experimentation. Conditions that permit detection of homologies between various unequally diverged regions of sequenced viral genomes have been investigated.[21] In the chorion system, we have noted that sequences that are highly divergent overall often show a smaller T_m depression than expected.[22] As an example, the B family clones pc401 and pc408 differ by 37.2% overall, but only by 13.6% in the conserved central domain[9]; the change in T_m of the cross hybrids, as determined from dot blots,[22] is approximately 10°.

It must be stressed again that the formulas given above were derived for solution hybridization, and thus are only rough approximations for filter hybridization. For example, it would not be surprising if the effective length of filter-bound hybrids is reduced by steric constraints. In agreement with this possibility, we have the impression that for filter self-hybridizations k declines more rapidly than expected from Fig. 1, when the temperature is raised to within 15° to 5° of the experimentally determined T_m. The T_m itself may be slightly depressed for filter-bound hy-

[18a] R. C. Yang, A. Young, and R. Wu, *J. Virol.* **34**, 416 (1980).
[19] R. B. Wallace, M. Schold, M. J. Johnson, P. Bembek, and K. Itakura, *Nucleic Acids Res.* **9**, 3647 (1981).
[20] S. Gillam, K. Waterman, and M. Smith, *Nucleic Acids Res.* **2**, 625 (1975).
[21] P. M. Howley, M. A. Israel, M.-F. Law, and M. A. Martin, *J. Biol. Chem.* **254**, 4876 (1979).
[22] G. K. Sim, F. C. Kafatos, C. W. Jones, M. D. Koehler, A. Efstratiadis, and T. Maniatis, *Cell* **18**, 1303 (1979).

brids. For example, for β-globin DNA–DNA dot hybrids of 51% G + C, 580 bp at 0.3 *M* Na^+ and 50% formamide, we have observed[3] a T_m of 56°, whereas the formula presented above would predict 60.6° for reversible T_m, and therefore at least 67° for irreversible T_m.

Two commonly overlooked variables in differential hybridization, which affect discrimination among nucleic acid sequences, are the extent of the reaction and the mass ratio of the radioactive probe to the unlabeled, filter-bound nucleic acid. In general, the hybridization of the probe will show the following kinetics when two (or more) filter-bound sequences react in the same bag with a probe in solution

$$dC/dt = -kC^2 - k_f IC - k_i H_i C \quad (1)$$

where C is the concentration of single-stranded probe at time t; k is the reassociation rate constant for the probe sequence in solution; k_f is the reassociation rate constant between the probe and the identical sequence, I, bound to the filter; and k_i is the reassociation rate constant between the probe and the cross-hybridizing sequence H_i, bound to the filter.

The kinetics will differ significantly, depending on whether the filter-bound nucleic acid or the probe is in excess, and in the latter case whether or not the probe can itself reassociate. For simplicity, let us assume that all filter-bound sequences are equal in mass ($I = H_i$), and that $k_f = k$. (In fact $k_f < k$, but the exact magnitude of the difference is not known, and it does not change the results qualitatively.) Then, if the filter-bound nucleic acid is in vast excess, for example, in typical dot blots,[3] the hybridization reactions follow pseudo-first-order kinetics; when they go to completion, the fraction of the probe hybridized to sequence i becomes

$$k_i \Big/ \sum_{i=1}^{m} k_i \quad (2)$$

where m is the number of filter-bound sequences.

At any time during the reaction, the ratio of the amount of probe hybridized to sequence i relative to sequence j is given by k_i/k_j (Fig. 2); i.e., the ratio is invariant with respect to time. In this case, discrimination between self- and cross-hybrids is not affected by the extent of the reaction (Fig. 2), because all the filter-bound sequences continuously compete for the same, limiting probe. (Discrimination will be affected if different samples are hybridized separately, but that situation does not normally apply.)

If the radioactive probe is in excess over the filter-bound sequences (e.g., typical Northern blots,[5] genomic Southern blots,[4] and initial screens of genomic or cDNA libraries), discrimination will generally decline as

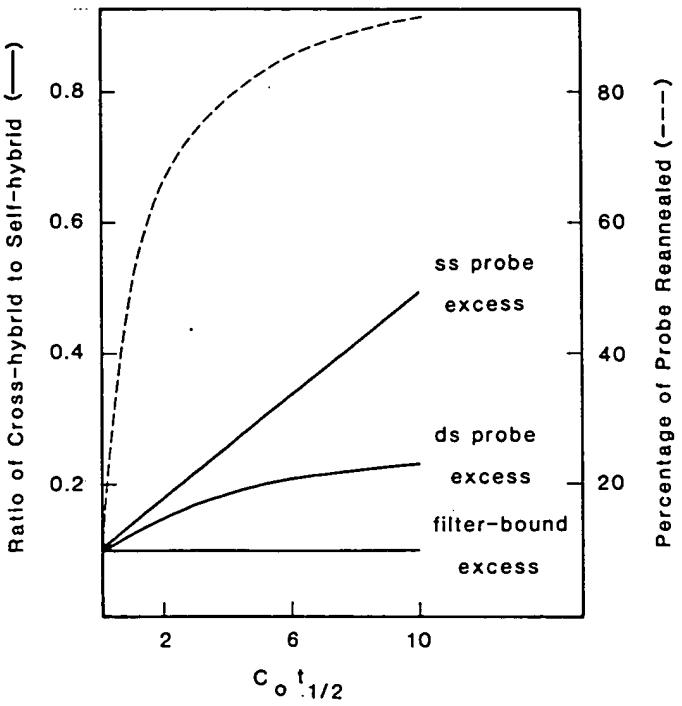


FIG. 2. Effective discrimination between self-hybrids and cross-hybrids, as a function of the extent of the reaction. The solid line shows the ratio of the amounts of probe hybridized to filter-bound heterologous and homologous sequences, $E_i(t)/E(t)$, for each of the three cases discussed in the text: filter-bound sequence in excess, double-stranded (ds), denatured probe in excess, and single-stranded (ss) (e.g., M13) probe in excess. The discrimination ratio, k_i/k, is assumed to be the same in all cases, 0.1. The dashed line shows normal C_0t kinetics of denatured, double-stranded DNA reannealing in solution.

the reaction proceeds. If the probe in solution can self-anneal, the relevant equation is

$$E_i(t)/E(t) = [1 - (1 + kC_0t)^{-n}]\,[1 - (1 + kC_0t)^{-1}]^{-1} \qquad (3)$$

where $E(t)$, $E_i(t)$ = extent of the reaction, i.e. the fraction of filter-bound DNA that has hybridized to the probe, for the self-hybrid and cross-hybrid i, respectively; C_0 = concentration of single-stranded probe at time 0; $n = k_i/k$, i.e., the discrimination ratio.

It can be seen that the effective discrimination, i.e., extent of cross-hybridization over self-hybridization, equals the discrimination ratio k_i/k very early in the reaction and deteriorates with the kinetics described above and as shown in Fig. 2.

If the probe in solution cannot self-anneal (e.g., M13 probes), the effective discrimination is given by

$$\frac{E_i(t)}{E(t)} = [1 - e^{-k_i C_0 t}]\,[1 - e^{-kC_0 t}]^{-1} \quad (4)$$

The parameters are defined above. Effective discrimination again equals k_i/k very early in the reaction, but deteriorates very rapidly (Fig. 2).

In summary, hybridizations should ideally be done for very short times: discrimination among various clones will then be maximal, and the same under conditions of either probe excess or filter-bound DNA excess. In practice, however, short hybridization times may not generate a sufficient autoradiographic signal. Thus, to distinguish among related sequences, it is best to use conditions of filter-bound DNA excess, as in typical dot blots[3] (see below).

A Strategy for Systematic Analysis of Multigene Families

Recovery of the Family

To begin the study of a multigene family for which a probe already exists, we recommend that a sublibrary be prepared by screening the appropriate library (cDNA or genomic) at a very permissive criterion [$k_i \simeq k$, $n = 1$ in Eq. (3)], using the Grunstein–Hogness[2,23] or Benton–Davis[1] procedures. At this stage we do not attempt to achieve discrimination, but only to recover as complete a collection of family members as possible; as discussed above, these two goals cannot be pursued simultaneously. A criterion of 30–35° below T_m and high salt (e.g., 0.6 *M* NaCl) for speeding up the hybridization reaction are recommended. To avoid unnecessary increase in the background, which is inherently high in the initial screening, the criterion should not be relaxed more than is necessary for the particular family. For completeness of the sublibrary, even very weakly hybridizing clones should be collected—especially since the plaque or colony size is frequently not uniform.

Such an initial screen will usually produce a mixture of true and false positives, which are distinguished by rescreening. Typical sublibraries that we have recovered in our study of chorion genes include from 200 to 600 clones, too many to make isolation of DNA from each clone practical. For sublibraries using bacterial plasmids as vectors, bacterial cultures are arranged in Microtiter plates, replica plated onto a nitrocellulose filter, placed on nutrient agar, and rescreened by the method of Grunstein and Hogness.[2] We have developed a similar replica plating method to rescreen

[23] M. Grunstein and J. Wallis, this series, Vol. 68, p. 379.

bacteriophage λ genomic sublibraries (see below). An example of the power and reproducibility of this method is shown in Fig. 3. Part a of Fig. 3 shows a rescreening of a genomic sublibrary under the same hybridization conditions as for the initial screen of the total library. This sublibrary had been selected to include genes closely homologous to the cDNA clone pc401, a B-family chorion sequence from *Antheraea polyphemus*.[9,22] APc110, a genomic clone containing two copies of chorion gene 401, and APc173, another genomic clone containing two copies of gene 10, a distantly related B-family sequence, were included as positive and negative controls.[9] A few blank spots were seen, corresponding to false positives of the initial screen, and a few clones grew poorly. Among the rest, some intensity differences were evident. However, even the negative control (APc173), showed significant hybridization under these conditions.

The next step is plaque or colony hybridization under stringent conditions. Figure 3b shows typical results, for the same sublibrary as in Fig. 3a. In this case, stringency was enhanced by a 15° increase in temperature and by use of a more specific probe. Instead of the complete pc401 DNA (565 bp), the probe contained a 257 bp fragment from the region corresponding to the 3′ end of pc401 mRNA. This probe encodes only a small portion of the conserved central protein domain, and consists largely of the more variable COOH-terminal arm sequence and nonconserved 3′ untranslated region.[9] Only 122 bp of this 3′-specific probe is well conserved within the B family (9% mismatch relative to pc10), whereas the entire pc401 probe includes a total of 255 bp of well conserved sequence. Clearly, the experiment of Fig. 3b achieved high discrimination. Hybridization with APc173 and with many of the clones in the sublibrary was essentially undetectable; only a few clones hybridized as intensely as APc110 itself, and others could be assigned to 3 or 4 classes on the basis of their degree of hybridization.

We perform plaque hybridizations to bacteriophage λ genomic sublibraries by the following procedure.

Preparation of Filters. Bacterial lawns are prepared by mixing 0.25 ml of 10 m*M* $MgCl_2$–10 m*M* $CaCl_2$, 0.5 ml of fresh bacterial overnight culture, and 6.5 ml of medium in 0.8% agarose. The mixture is poured over bottom agar (medium with 1.2% agar) in 150-mm petri dishes. Care should be taken to avoid air bubbles. We find that the bottom agar should be poured at least 2 days in advance to avoid condensation during phage growth. Allow the top agar mixture to gel for 10 min at room temperature. Using a replica plater,[24] transfer a small aliquot of each phage stock from

[24] M. Brenner, D. Tisdale, and W. F. Loomis, *Exp. Cell Res.* **90**, 249 (1975).

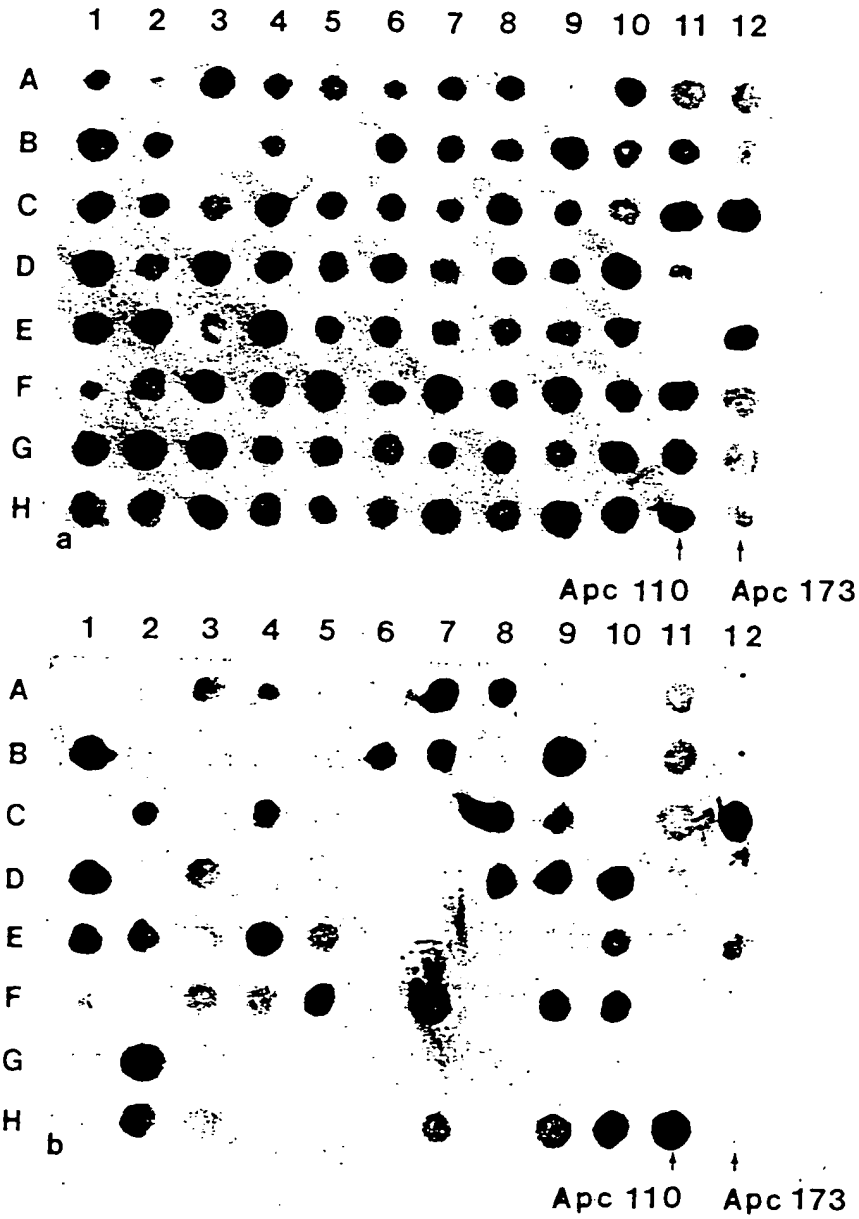


FIG. 3. An *Antheraea polyphemus* genomic DNA library was screened as described by Benton and Davis[1] with a combination of two chorion cDNA probes, pc18 and pc401. (The genes encoding chorion protein 18 are members of the A family and are each paired with a B-family gene encoding chorion protein 401.) Plaque-purified phage stocks from all positive clones were prepared. Aliquots of each stock were delivered into Microtiter wells and plated

Microtiter plates to the bacterial lawn. Allow the liquid to be absorbed into the agar (about 30 min), invert the plates, and incubate overnight at 37°. Uniform plaques of about 6 mm should develop. Chill the plates at 4° for 10–30 min before placing a sheet of dry nitrocellulose directly on the plaque-containing lawn. Precut nitrocellulose circles are commercially available, but we find it more economical and convenient to use rectangular sheets cut to 8.5 × 12.5 cm. Allow the DNA to transfer for 60 to 90 sec, remove the filter from the plate, and denature and fix the DNA to the nitrocellulose by soaking the filter in 0.1 *M* NaOH, 1.5 *M* NaCl for about 30 sec. Neutralize the filter by soaking in 0.5 *M* Tris-HCl, pH 7.5, 0.5 *M* NaCl for about 30 sec. Blot the excess buffer on 3 MM paper, air-dry, and bake under vacuum for 2 hr at 80°. We find that the same plate can be used to prepare at least four filters with no apparent loss in signal.

Prehybridization. Filters are prehybridized in 2–4× SET buffer (1× SET buffer is 0.15 *M* NaCl, 0.03 *M* Tris-HCl, pH 8, 1 m*M* EDTA), 10× Denhardt's solution, 0.2% SDS, and 50 μg/ml herring sperm DNA for at least 3 hr at the hybridization temperature.

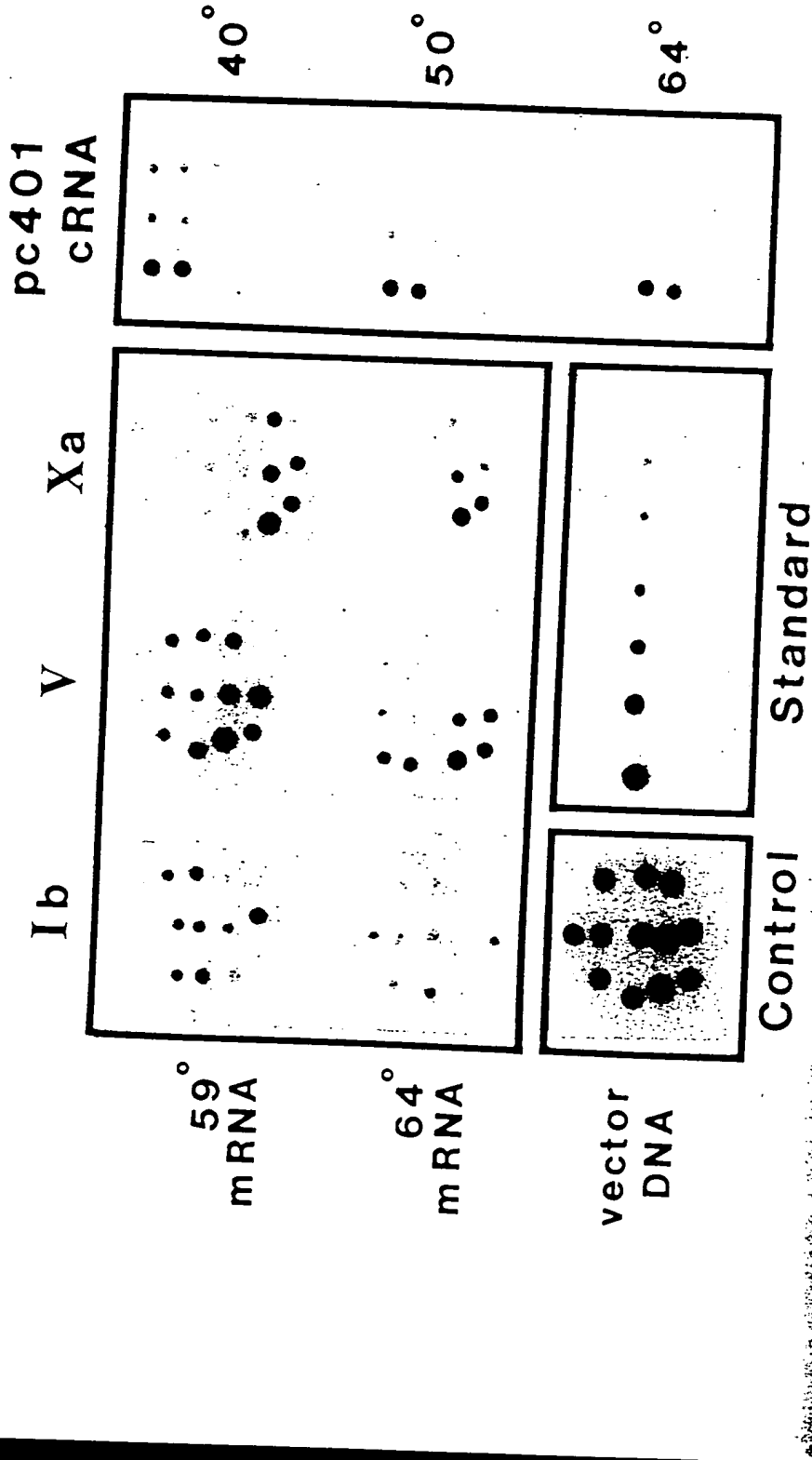
Hybridization and Washing. Filters are hybridized in prehybridization mixture supplemented with the radioactive probe and dextran sulfate to a final concentration of 10%. Although we normally hybridize overnight (12–16 hr), we find that the signals are sufficiently strong after hybridizing for only 5 hr. Filters are washed for 20 min each in two changes of 2× SET, 0.2% SDS and two changes of 1× SET, 0.1% SDS at the hybridization temperature.

Discrimination by Dot Hybridization

Once experiments such as those of Fig. 3b have focused attention on a limited number of clones, more refined analysis can be undertaken by dot-blot hybridization. The original paper describing the method[3] should be consulted, as well as a more recent review.[25] In this procedure, DNAs are purified from the clones of interest and immobilized on filters in precisely

[25] F. C. Kafatos, G. T. Thireos, C. W. Jones, S. G. Tsitilou, and K. Iatrou, *in* "Gene Amplification and Analysis," Vol. 2: "Structural Analysis of Nucleic Acids" (J. G. Chirikjian and T. P. Papas, eds.), p. 537. Elsevier/North-Holland, Amsterdam, 1981.

onto a lawn of bacteria. DNAs from the resulting plaques were transferred to nitrocellulose and hybridized to: (*a*) nick-translated pc401, at 65°, 0.6 *M* NaCl; (*b*) a nick-translated *Kpn*I fragment of pc401 representing the 3′ half of the mRNA, at 80°, 0.3 *M* NaCl. In each case hybridized filters were washed at the hybridization temperature in 0.3 *M* NaCl and 0.15 *M* NaCl. APc110 (a genomic clone containing two copies of 401) and APc173 (a genomic clone containing two copies of the distantly related, B-family chorion gene 10) were used as positive and negative controls.


40°
50°
64°
pc401 cRNA
Ib
V
Xa
Standard
Control
59° mRNA
64° mRNA
vector DNA

equal amounts and in dots of uniform diameter. As a result, after hybridization the intensity of the various dots can be estimated semiquantitatively, by visual comparison to standards consisting of a dilution series of radioactive DNA directly spotted on a similar filter.

Figure 4 shows examples of dot-blot hybridizations.[22] The panel labeled "Standard" is a 2-fold dilution series that spans a 64-fold range of radioactivity. The panel on the right exemplifies the discrimination that can be achieved between homologs by judicious choice of conditions. The experiment was performed in 50% formamide, and pc401 cRNA was used as the probe. At 40°, all seven members of the B family included in this filter hybridized, albeit to different extents; five clones of the A family were also present in the filter, but did not detectably hybridize. At 50° and 64° two dots hybridized intensely and almost equally; one was pc401 itself (top) and the other pc602, which differs from pc401 by less than 1% mismatching. Weak hybridization with pc10 was observed at 50° (about 10-fold less than the self-hybrid), but none at 64°.

The remainder of Fig. 4 exemplifies a different, widely applicable use of dot-blot hybridization: assaying the concentrations of various sequences in a series of samples. Such assays are enormously more convenient and more sensitive than assays by liquid hybridization and, by careful attention to the conditions, are only slightly less accurate. The example shown permitted classification of 12 chorion cDNA clones into "early," "middle," and "late" developmental classes, depending on their preferential hybridization to mRNA preparations from the corresponding developmental stages (Ib, V, and Xa, respectively). In this case, the cloned DNAs were fixed to the filter and the labeled probes were mRNAs containing these sequences in unknown amounts. It is also possible to attach the unknown samples (either DNA or RNA) to the filter and probe them with a specific cloned sequence; examples of this approach are given by

FIG. 4. Discrimination of homologous sequences by dot hybridization.[22] *Top left:* Changing concentrations of specific chorion RNA sequences during development. Poly(A)-containing cytoplasmic RNA from *Antheraea polyphemus* follicles at stages Ib, V, or Xa was isolated, end-labeled with ^{32}P after alkali treatment, and hybridized to replicate filters containing equal amounts of spotted DNA from 12 different chorion cDNA clones. Hybridizations were performed in 50% formamide, 0.6 *M* NaCl for 28 hr (59°) or 40 hr (64°). As a control (*lower left*) one filter was hybridized with ^{32}P labeled pML-21 DNA, the vector used in cloning. *Right:* Cross-hybridization of chorion cDNAs. DNAs from 12 chorion cDNA clones (7 from the B family and 5 from the A family) were again spotted on nitrocellulose. ^{32}P-labeled 401 cRNA was prepared using *E. coli* RNA polymerase and the nuclease S1 excised insert of pc401. Hybridizations were performed in 50% formamide, 0.6 *M* NaCl at the temperatures indicated. *Standard:* A two-fold dilution series of [^{32}P]DNA spotted directly on nitrocellulose.

Weisbrod and Weintraub.[26] In the case of multigene families, it is important to distinguish the concentrations of identical vs related sequences. In Fig. 4, this was accomplished by performing the experiment at both 59° and 64°.

Exactly the same approach can be used to characterize members of the multigene families by hybrid-selected translation[6]: one hybridizes DNA dots under discriminating conditions with "probe" consisting of unlabeled mRNA and then melts the hybrids and translates the RNA in a cell-free system by standard procedures. If hybridization is performed under conditions that permit formation of some cross-hybrids, this technique can be used to detect the degree of homology between protein species for which sequence information is not available.[27,28]

An apparatus is now available from Bethesda Research Laboratories and Schleicher & Schuell for performing dot blots with multiple samples stored in Microtiter plates. Dot blots can also be performed manually, and we recommend the following procedure.[25]

Preparation of Filters. All filters are washed in water for 1 hr. A plain nitrocellulose filter (22 mm in diameter) is mounted on top of two nitrocellulose filters with grids, on a sintered-glass platform connected to a water aspirator. The filters are washed 3–4 times with 1 *M* ammonium acetate before spotting with DNA.

Plasmid DNA is linearized by restriction endonuclease treatment and digested with proteinase K (200 μg/ml in 50 m*M* Tris-HCl, pH 7.5; at 37°, 30 min each with 0.2% and 2% SDS). After phenol extraction the DNA is denatured in 0.3 *N* NaOH for 10 min and chilled; when needed, it is diluted with an equal volume of cold 2 *M* ammonium acetate to a concentration of 1.4 μg/ml. It is then taken up in a capillary pipette attached to a micropipette filler (Clay Adams suction apparatus No. 4555), and is spotted on the filter under light vacuum; once the pipette touches the filter, contact is maintained continuously, while the DNA solution is delivered slowly with a combination of vacuum suction and positive pressure from the filler. We routinely spot 0.7 μg of DNA in 50 μl per dot, using a 100-μl micropipette. The dot is washed with a drop of 1 *M* ammonium acetate, as is the area to be spotted next. After all the dots are made, the filter is washed again with 1 *M* ammonium acetate under suction, air dried, treated with 2× Denhardt's solution for 1 hr, drained, air-dried, and baked at 80° for 2 hr.

Prehybridization. The dried filters are wetted evenly by slow immersion in 10× Denhardt's–4× SET buffer and are shaken in that solution for

[26] S. Weisbrod and H. Weintraub, *Cell* 23, 391 (1981).
[27] N. K. Moschonas, Ph.D. Thesis, Univ. of Athens, Athens, Greece, 1980.
[28] G. Thireos and F. C. Kafatos, *Dev. Biol.* 78, 36 (1980).

at least 1 hr. They are transferred to sterile, siliconized scintillation vials containing blank hybridization mixture, e.g., 50% deionized formamide, 2× Denhardt's solution, 4× SET, 0.1% SDS, 100 μg of yeast tRNA and 125 μg of poly(A) per milliliter, and are incubated for at least 1 hr at the hybridization temperature.

Hybridization and Washing. The filters are hybridized for 16–48 hr in hybridization mixture (prehybridization mixture supplemented with radioactive probe). The temperature is selected to give the desired criterion of stringency. Washes are performed at the same temperature, twice each with 4×, 2×, 1×, 0.4×, and 0.2× SET, all with 0.1% SDS. Two final washes are performed at room temperature, with 0.1× SET without SDS. The filters are covered with Saran wrap and autoradiographed, either dry or moist (if melts are to be undertaken).

Discrimination by Southern and Northern Analysis

Discrimination between members of a multigene family, by careful control of hybridization conditions and selection of appropriately specific probes, can also be accomplished in Southern and Northern transfer experiments. While dot hybridization is preferable for quantitation and is somewhat more convenient, Northern analyses permit detection of the sizes of the hybridizing transcripts (which may vary during development), and Southern experiments using genomic DNA clones can identify restriction fragments that contain the cross-hybridizing sequences. Thus, these three filter hybridization procedures are complementary, and all are apt to be used in the study of a multigene family.

Figure 5 shows an example of discrimination by low- and high-stringency Southerns. The probe in this case was the high-cysteine cDNA clone m2574 of *Bombyx mori*,[29] and the filter-bound DNAs were derived from 18 overlapping genomic clones (B1 to B18) totaling 270 kb of continuous DNA spanning the high-cysteine region of the chorion locus.[30] Fifteen m2574-like genes were detected by low-stringency hybridization (75°, 0.6 *M* NaCl). High-stringency hybridization permitted identification of gene 12 as the one from which m2574 was derived, genes 14 and 15 as very close homologs, genes 8 and 10 as moderately close homologs, and the remaining 10 genes as more distant homologs. In this case, high stringency was accomplished by high temperature (85°, 0.3 *M* NaCl) and by the use of a relatively high concentration of DNase during nick-translation of the probe. The high DNase approach is recommended for maximizing

[29] K. Iatrou, S. G. Tsitilou, and F. C. Kafatos, *J. Mol. Biol.* 157, 417 (1982).
[30] T. H. Eickbush and F. C. Kafatos, *Cell* 29, 633 (1982).

FIG. 5. Identification of a subset of related genes within a multigene family. Approximately 1-μg aliquots of DNA from genomic clones B1 through B18, which span the high-cysteine region of the chorion locus, were digested with *Eco*R1, electrophoresed through an agarose gel, transferred to nitrocellulose paper by the Southern method,[4] and hybridized with nick-translated cDNA probe m2574. The low-stringency hybridization (*left*) was at 75°, 0.6 *M* NaCl with the probe nick-translated in the presence of 5 ng of DNase I per milliliter. The high-stringency hybridization (*right*) was at 85°, 0.3 *M* NaCl, with the probe nick-translated in the presence of 125 ng of DNase I per milliliter. The numbers under each hybridized fragment refer to the m2574-like genes as numbered in Eickbush and Kafatos.[30]

discrimination when sequence information is not available, and therefore specific probes cannot be tailored as in Fig. 3b. High DNase results in short probe lengths, and, as a result, if mismatching is nonrandomly distributed, only the short fragments that might contain a conserved sequence will cross-hybridize at stringent conditions. By contrast, with long probes, signals from hybrids that are stable because of a short conserved sequence will be enhanced by the presence of covalently attached unpaired tails, resulting in decreased discrimination.

In conclusion, deliberate control of the stringency of hybridization is possible in all filter hybridization procedures and is a powerful tool in the isolation and characterization of multigene families.

Acknowledgments

Original work was supported by grants from NIH, NSF, and the American Cancer Society.

nitrocellulose paper by the Southern method,[4] and hybridized with nick-translated cDNA probe m2574. The low-stringency hybridization (*left*) was at 75°, 0.6 *M* NaCl with the probe nick-translated in the presence of 5 ng of DNase I per milliliter. The high-stringency hybridization (*right*) was at 85°, 0.3 *M* NaCl, with the probe nick-translated in the presence of 125 ng of DNase I per milliliter. The numbers under each hybridized fragment refer to the m2574-like genes as numbered in Eickbush and Kafatos.[30]

[20] Synthesis of ds-cDNA Involving Addition of dCMP Tails to Allow Cloning of 5′-Terminal mRNA Sequences

By HARTMUT LAND, MANUEL GREZ, HANSJÖRG HAUSER, WERNER LINDENMAIER, and GÜNTHER SCHÜTZ

Cloning of mRNA sequences after reverse transcription into cDNA copies[1-3] plays an important role in the analysis of gene structure and function. A variety of methods for cloning cDNAs in bacterial plasmids has been described.[4-10] In the method most commonly used,[7-9] the ability

[1] I. Verma, G. F. Temple, H. Fan, and D. Baltimore, *Nature (London) New Biol.* **235**, 163 (1972).
[2] D. L. Kacian, S. Spiegelman, A. Bank, M. Terada, S. Metafora, L. Dow, and P. A. Marks, *Nature (London) New Biol.* **235**, 167 (1972).
[3] J. Ross, H. Aviv, E. Scolnick, and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 264 (1972).
[4] F. Rougeon, P. Kourilsky, and B. Mach, *Nucleic Acids Res.* **2**, 2365 (1975).
[5] T. H. Rabbits, *Nature (London)* **260**, 221 (1976).
[6] K. O. Wood and J. C. Lee, *Nucleic Acids Res.* **3**, 1961 (1976).
[7] T. Maniatis, S. G. Kee, A. Efstratiadis, and F. C. Kafatos, *Cell* **8**, 163 (1976).
[8] R. Higuchi, G. V. Paddock, R. Wall, and W. Salser, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3146 (1976).
[9] F. Rougeon and B. Mach, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3418 (1976).
[10] S. Zain, J. Sambrook, R. J. Roberts, W. Keller, M. Fried, and A. R. Dunn, *Cell* **16**, 851 (1979).

ISBN 0-12-182000-9